

## BASIC RESEARCH

# Articles with short titles describing the results are cited more often

Carlos Eduardo Paiva,<sup>1,II</sup> João Paulo da Silveira Nogueira Lima,<sup>I</sup> Bianca Sakamoto Ribeiro Paiva<sup>II</sup>

<sup>I</sup>Barretos Cancer Hospital, Department of Clinical Oncology, Division of Breast and Gynecological Cancers, Barretos/SP, Brazil. <sup>II</sup>Barretos Cancer Hospital, Post-graduation Program, Barretos/SP, Brazil.

**OBJECTIVE:** The aim of this study was to evaluate some features of article titles from open access journals and to assess the possible impact of these titles on predicting the number of article views and citations.

**METHODS:** Research articles (n=423, published in October 2008) from all Public Library of Science (PLOS) journals and from 12 Biomed Central (BMC) journals were evaluated. Publication metrics (views and citations) were analyzed in December 2011. The titles were classified according to their contents, namely methods-describing titles and results-describing titles. The number of title characters, title typology, the use of a question mark, reference to a specific geographical region, and the use of a colon or a hyphen separating different ideas within a sentence were analyzed to identify predictors of views and citations. A logistic regression model was used to identify independent title characteristics that could predict citation rates.

**RESULTS:** Short-titled articles had higher viewing and citation rates than those with longer titles. Titles containing a question mark, containing a reference to a specific geographical region, and that used a colon or a hyphen were associated with a lower number of citations. Articles with results-describing titles were cited more often than those with methods-describing titles. After multivariate analysis, only a low number of characters and title typology remained as predictors of the number of citations.

**CONCLUSIONS:** Some features of article titles can help predict the number of article views and citation counts. Short titles presenting results or conclusions were independently associated with higher citation counts. The findings presented here could be used by authors, reviewers, and editors to maximize the impact of articles in the scientific community.

**KEYWORDS:** Articles; Citations; Visualization; Titles.

Paiva CE, da Silveira Nogueira Lima JP, Paiva BS. Articles with short titles describing the results are cited more often. Clinics. 2012;67(5):509-513.

Received for publication on March 5, 2012; First review completed March 29, 2012; Accepted for publication on March 29, 2012

E-mail: caredupai@gmail.com

Tel.: 55 17 3321-6600

## INTRODUCTION

Citation rates are used to measure the impact of articles, journals, and even researchers. The most well-known and established rate is the journal impact factor (JIF), released by *Journal Citation Reports* (JCR), which evaluates thousands of journals using citation data. In addition to the JIF, the *Journal of Citation Reports* offers a variety of impact and influence metrics (1). Other citation databases have become available, such as Scopus (2) and Google Scholar (3). Despite severe criticism of the limitations and biases of the JIF, this method has been consolidated as the single most important scientific production metric tool.

To increase the visibility of their research, researchers want to have their work published in high-impact journals. Publishing manuscripts with high citation potential is also

of interest to scientific journals, as doing so can improve the journal's credibility, relevance, and financial independence. In this regard, it seems to be very important to identify the manuscript characteristics associated with a higher number of citations, as well as more views from journal readers.

The article's title has the challenging task of triggering the curiosity of readers by inviting them to appraise the article and perhaps use it as a reference for new research. Thus, the title is the most important summary of a scientific article. It is generally the first (and sometimes the only) information obtained from the published article.

Despite this theoretical importance of titles, the recommendations of scientific journal editors regarding article titles are largely based on their personal experiences. With regard to biomedical journals, only two published studies (4-5) have evaluated article titles to identify features that could predict the number of subsequent citations of a published article. Despite the publication of previous studies evaluating the role of title features on scientific relevance, little is known about articles published in open access journals. Some of these open journals were created in attempts to circumvent problems in knowledge dissemination.

**Copyright** © 2012 CLINICS – This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

No potential conflict of interest was reported.

The aim of the present study were to evaluate some features of article titles from open access journals, to determine the existence of any relationship between the article title and its relevant dissemination, and to associate the title with the number of article views and citations.

## MATERIALS AND METHODS

### Selection of journals and articles

During the journal selection process, we sought to obtain a sizable number of biomedical articles with available citation and page view information. Therefore, open access journals from the *BioMed Central* (BMC) and *Public Library of Science* (PLOS) publishing groups were gathered to form the present database. All six PLoS journals, as well as the six best ranked and the six worst ranked BMC journals, according to JCR 2010, were included in the analysis (Table 1).

All original research articles published from September 1, 2008, to September 31, 2008, were analyzed. Articles classified as review articles, case reports, commentaries, editorials, and letters to the editor were excluded from the analysis. The one-month-only period of inclusion was justified based on the premise that articles published earlier would have had longer exposure, allowing for more citations by others, compared to articles that were published later with a shorter "reading time." The three-year period spanning from the article publication to the present analysis was considered to be a sufficient amount of time to measure the impact of a specific article in the scientific community.

### Metrics extraction

The numbers of times the article was viewed at the publisher site, downloaded, and cited according to JCR Science Edition 2010 were collected for the period from December 6, 2011, to December 20, 2011.

**Table 1** - Selected journals with their respective numbers of articles analyzed and impact factors.

| Journal                                    | N   | IF*    |
|--|-----|--------|
| <i>PLOS group</i>                          |     |        |
| PLOS Biology                               | 18  | 12.472 |
| PLOS Medicine                              | 5   | 15.617 |
| PLOS Computational Biology                 | 19  | 5.515  |
| PLOS Genetics                              | 29  | 9.543  |
| PLOS Pathogens                             | 22  | 9.079  |
| PLOS One                                   | 190 | 4.411  |
| PLOS Neglected Tropical Diseases           | 13  | 4.752  |
| <i>Biomed Central</i>                      |     |        |
| BMC Medicine <sup>a</sup>                  | 1   | 5.750  |
| BMC Biology <sup>a</sup>                   | 4   | 5.203  |
| BMC Genomics <sup>a</sup>                  | 37  | 4.206  |
| BMC Plant Biology <sup>a</sup>             | 9   | 4.085  |
| BMC Medical Genomics <sup>a</sup>          | 9   | 3.766  |
| BMC Evolutionary Biology <sup>a</sup>      | 22  | 3.702  |
| BMC Musculoskeletal Disorders <sup>b</sup> | 11  | 1.941  |
| BMC Pediatrics <sup>b</sup>                | 5   | 1.904  |
| BMC Health Services Research <sup>b</sup>  | 18  | 1.721  |
| BMC Family Practice <sup>b</sup>           | 6   | 1.467  |
| BMC Ophthalmology <sup>b</sup>             | 2   | 1.375  |
| BMC Medical Education <sup>b</sup>         | 3   | 1.201  |

\*Impact factor (IF) according to JCR Science Edition 2010. <sup>a</sup> Higher cited group from BMC. <sup>b</sup> Lower cited group from BMC.

A pre-defined form was used to collect the article features. Relevant items extracted from the article titles included the number of characters, the use of question marks, reference to a geographical area (city, state, and country), and the use of a hyphen or colon separating different ideas within a sentence. Two authors independently analyzed the titles to classify them into three distinct categories: type 1, articles describing the research methods/design (methods-describing title); type 2, articles describing the results/conclusions (results-describing title); and type 3, articles that were non-classifiable. In the case of classification disagreements, the authors tried to reach a final consensus. The numbers of characters in the titles were divided into three different groups according to percentiles 25 (P25) and 75 (P75), i.e.,  $\leq P25$ , between P25 and P75, and  $> P75$ .

### Statistical analysis

The data are presented as medians and interquartile ranges (IQRs). The comparisons between article title features and visibility were performed using the nonparametric Mann-Whitney U test and Kruskal-Wallis test, followed by Dunn's multiple comparison post test. Spearman's coefficient (r) test was used to investigate the relationship between the number of characters in the title and the view and citation counts.

A stepwise linear regression model was used to evaluate the independent variables that predicted citation rates. The covariates that were utilized in the multivariate model were as follows: number of characters (continuum variable), type of article title (1 vs. 2), use of question marks (yes vs. no), reference to a geographical area (yes vs. no), and use of a hyphen or colon to separate different ideas within a sentence (yes vs. no).

The statistical analyses were performed using GraphPad Prism3 (San Diego, CA, USA). A *p*-value less than 0.05 was considered statistically significant.

## RESULTS

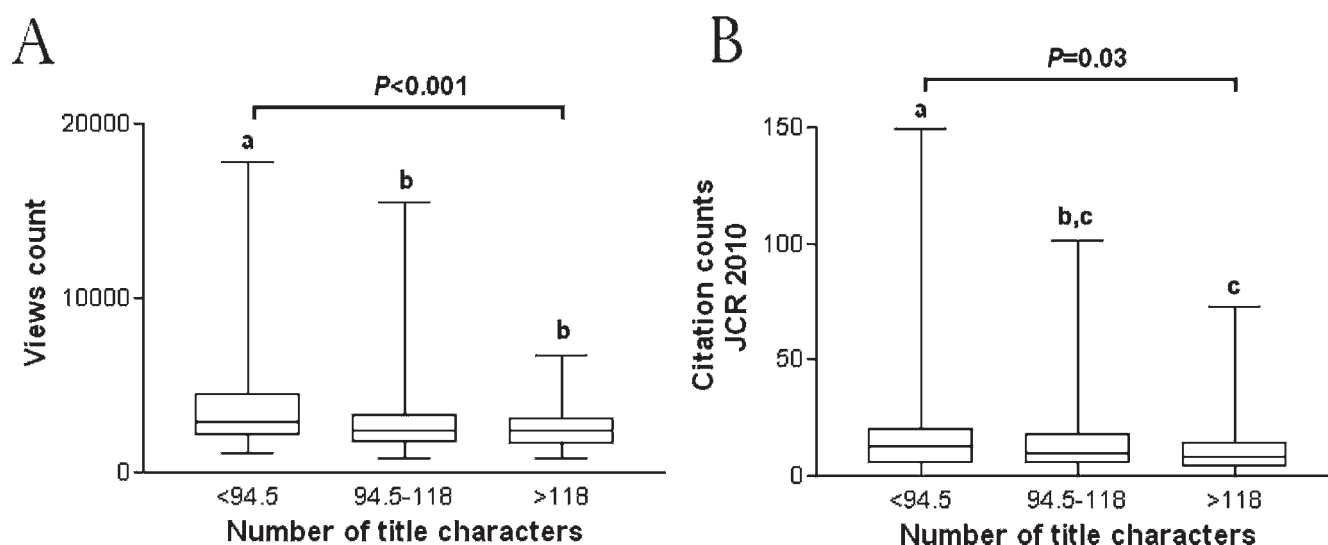
In total, 423 original research article titles were included in the analysis; the article distribution, according to journal, is described in Table 1.

The median (IQR) number of views and citations were 2533 (1744) and 10 (13), respectively. There was a positive correlation between the number of views and citations ( $r = 0.434$ ,  $p < 0.001$ ). The median (IQR) number of title characters was 94 (43.5).

There were weak and negative correlations between the number of characters in the title and the numbers of article views and citations ( $r = -0.168$ ,  $p < 0.001$  and  $r = -0.104$ ,  $p = 0.032$ , respectively).

The median (IQR) numbers of views, according to the number of title characters, were 2892 (2404), 2446 (1655), and 2359 (1439) for the groups of article titles with  $\leq 94.5$  characters, 94.5 to 118 characters, and more than 118 characters, respectively ( $p < 0.001$ ). The group with the fewest characters ( $\leq 94.5$ ) had significantly more views compared to the other two groups based on the post test analysis ( $p < 0.01$  for both) (Figure 1A).

Regarding citation rates, the median (IQR) numbers of citations were 12.5 (15), 10 (13), and 8 (10) for the groups with  $< 94.5$  characters, 94.5 to 118 characters, and more than 118 characters, respectively ( $p = 0.034$ ). Post-hoc analysis showed that the group with  $< 94.5$  characters had more citations than the group with  $> 118$  characters ( $p < 0.05$ ; Figure 1B).



**Figure 1** - View and citation counts according to the numbers of characters in the titles. A) The numbers of views were statistically different among the three groups analyzed ( $p < 0.001$ ). Post-hoc analyses showed that the group with the least number of characters ( $\leq 94.5$ ) had significantly higher view counts compared with the other two groups (94.5 to 118 and  $> 118$ ) ( $p < 0.01$  for both). B) Citation counts were statistically significantly different among the three groups analyzed ( $p = 0.034$ ). Post-hoc analyses showed that the group with the least number of characters ( $\leq 94.5$ ) had significantly higher view counts compared with the group with the greatest number of characters ( $> 118$ ) ( $p < 0.05$ ). Different letters (a, b, and c) designate statistically significant group differences.

There were 231 (54.6%) methods-describing titles (type 1), 171 (40.4%) results-describing titles (type 2), and 21 (4.9%) non-classifiable titles (type 3). The median numbers of views were not different between groups of articles with different typologies ( $p = 0.111$ , data not shown). In contrast, the median number (IQR) of citations for type 1 articles was 8 (10.5), which was significantly less than the median number of citations for type 2 articles (median = 12, IQR = 13) ( $p < 0.001$ ; Figure 2A).

The presence of a question mark in the title had no impact on the viewing rate ( $p = 0.782$ , data not shown). The median number of citations was lower in article titles containing question marks ( $n = 11$ , median = 6) compared with article titles without question marks ( $n = 412$ , median = 10) ( $p = 0.046$ ; Figure 2B).

Regarding the number of views, there was no difference between the groups of titles either describing or not describing a geographic location ( $p = 0.906$ , data not shown). Titles referring to a specific geographical region were significantly less cited ( $n = 35$ , median = 5) than titles that did not reference a specific region ( $n = 388$ , median = 10) ( $p < 0.001$ ; Figure 2C).

Article titles with two components separated by a colon or a hyphen ( $n = 93$ , median = 7) had fewer citations compared with titles that did not include these components ( $n = 330$ , median = 10) ( $p = 0.004$ ; Figure 2D). Regarding the number of article views, there was no difference between the groups ( $p = 0.427$ , data not shown).

The results of the linear regression analyses showed that only article title typology (beta coefficient = 5.458, standard error = 1.601,  $t = 3.409$ ,  $p = 0.001$ ) and the number of title characters (beta coefficient = -0.066, standard error = 0.027,  $t = -2.445$ ,  $p = 0.015$ ) were statistically significant predictors of citation rates in the final model ( $F = 7.581$ ,  $p = 0.001$ ).

## DISCUSSION

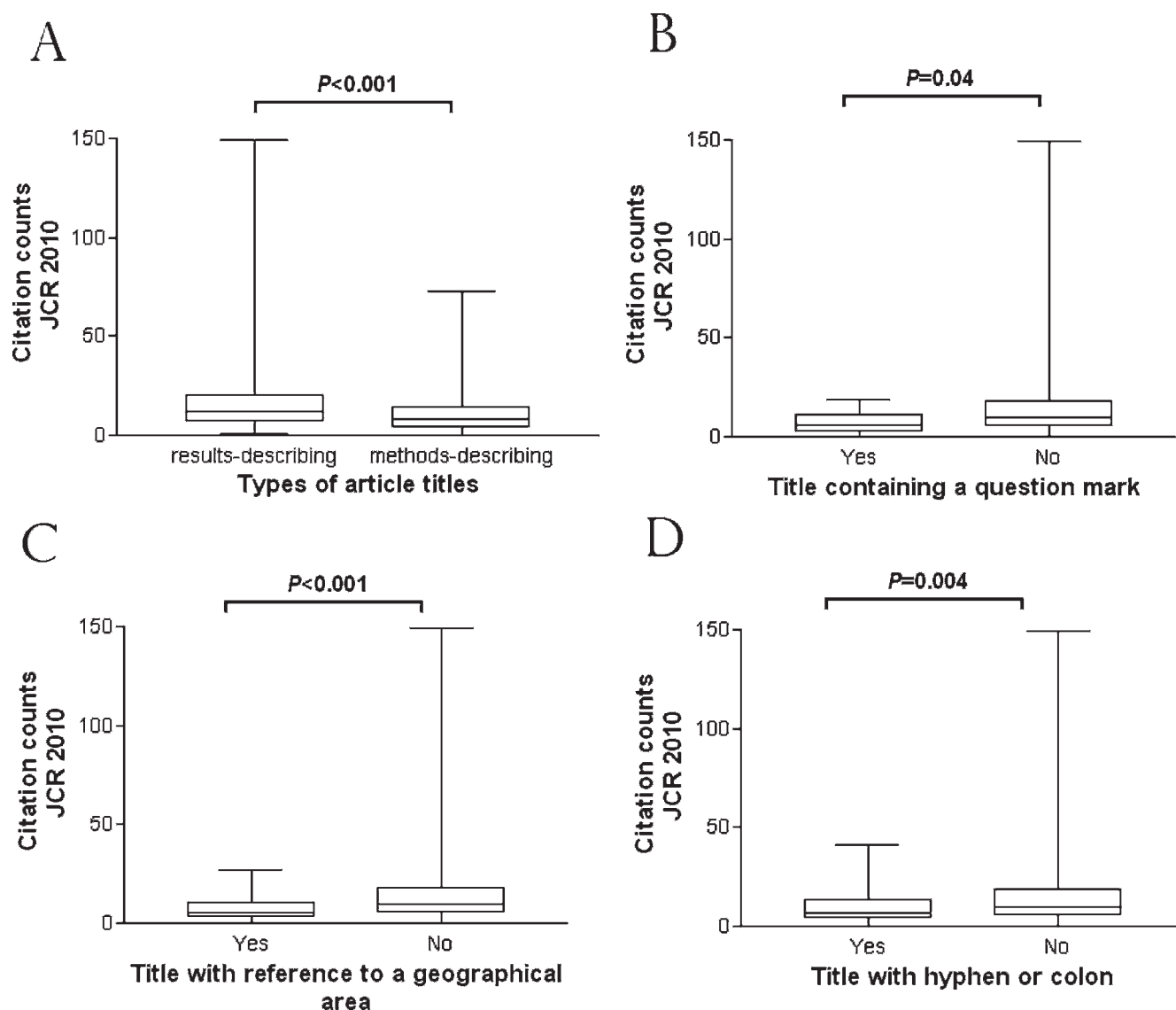
The present study addressed the association of textual features of scientific article titles with the articles' visibility in the scientific media. The study's findings highlight the relevance of analyzing title features during the pre-publication process.

Journal editors and experienced authors frequently suggest the use of a short, concise, and informative title (6-8). Some scientific journals impose a maximum limit on the number of words or characters in titles (9-10); however, such editorial guidelines are not based on scientific data.

Shot-titled articles might be more attractive to readers than articles with longer titles; the latter could be seen as complex or boring (8). If readers cannot understand a title, there is only a small chance that they will read the abstract or the full paper (6). In this regard, a negative correlation would be expected between the number of characters in an article title and the number of article views, which was indeed confirmed in the present study, despite the small rho value found.

The relevance of the new electronic methods of knowledge dissemination investigated in this study, namely article viewing and article download, has become increasingly recognized. To our knowledge, no published research studies have addressed the effect of article title length on the number of views.

Currently, literature searches are carried out by electronic means based on online database searches. For instance, several medical groups have developed electronic research methods to improve and optimize article retrieval. Other than these professional search methods, the overwhelming majority of searches are restricted to title or keyword searches only. Therefore, titles containing more words/characters should have a higher probability of being found using such searching strategies. In this regard, two different published studies found that longer article titles received



**Figure 2** - Citation counts according to some features of article titles. A) Articles with results-describing titles were cited more often than those with methods-describing titles ( $p < 0.001$ ). B) Articles with titles containing a question mark were cited less often than those without such punctuation ( $p = 0.046$ ). C) Articles with titles referring to a specific geographic region were cited significantly less often than those without reference to a specific region ( $p < 0.001$ ). D) Articles with titles containing two components separated by a colon or a hyphen had a lower number of citations compared to those with titles without this grammatical structure ( $p = 0.004$ ).

more citations (4-5). Titles are even more relevant to readers when selecting which articles will be used among those retrieved from journals' tables of contents, from searched databases, and from scanned bibliographies. In contrast, the present study showed that short titles have a higher probability of being cited by other papers. It is hypothesized that, at least in open access journals, shorter-titled articles are cited more often because they are viewed more often.

The *British Medical Journal* recommends that titles include the study design if the paper presents original research (11). In fact, 96% of articles published in the *BMJ* during 2001 could be classified as having titles of the methods-describing type (12). In the present study, article titles summarizing results or conclusions were associated with higher citation rates compared with methods-describing titles. Ultimately, what readers really want to know about a paper is its main results. The findings of the present study could be hypothesis-generating,

forming evidence to be considered by future authors, reviewers, and journal editors.

Our findings are in agreement with those of other authors who showed that titles with references to specific geographical regions were associated with fewer citations (4). This finding probably limits the visibility of an article to specific readers.

Earlier studies that addressed title features with regard to citation metrics used different designs (4-5). In particular, they compared title characteristics between the most cited and least cited articles. The present analysis seems to be more realistic because we systematically studied all of the published research articles during a defined period of time.

Regarding the use of a colon or a hyphen to separate two distinct components of a title, our findings are in accordance with expert opinion (6), suggesting that authors should avoid such punctuation. In contrast, the most cited articles

had a greater number of titles containing a colon compared with the least cited articles (4).

Multivariate analysis was performed to evaluate the title features that could predict citation rates. Titles with a smaller number of characters and those describing results were cited more often. To our knowledge, this is the first study to evaluate article title features from open access journals as predictors of citation rates.

Our study has some limitations. First, only a group of journals and their articles were analyzed over a specific period of time. The articles sampled might not represent those of all biomedical journals. Another limitation of this study is that it analyzed only features from article titles, although other parts of manuscripts are obviously of great importance, such as their scientific content.

In conclusion, some features of article titles can be used to predict the numbers of views and citations of articles. Articles with short titles are more often viewed and cited by others. Articles with titles containing a question mark, with references to specific geographical regions, and with a colon or a hyphen were cited less often, especially compared to articles with titles summarizing research results or conclusions, which were cited more often. Based on the multivariate analysis, only short titles presenting results or conclusions were independently associated with higher citation rates. The findings presented here could be used by authors, reviewers, and editors to maximize the impact of articles in the scientific community.

## ACKNOWLEDGMENTS

The authors would like to thank the Learning and Research Institute of Barretos Cancer Hospital (Barretos, São Paulo, Brazil) for revising the English text.

## AUTHOR CONTRIBUTIONS

Paiva CE designed the study and was also responsible for the data collection, statistical analysis and manuscript draft. Lima JP designed the study and was also responsible for the manuscript draft. Paiva BS was responsible for the data collection and manuscript draft. All authors have read and approved the final manuscript.

## REFERENCES

1. Thomson Reuters. ISI Web of Knowledge Web site. 2011; Available at <http://wokinfo.com/>. Accessed December 13, 2011.
2. Elsevier. Scopus Web site. 2011; Available at <http://www.scopus.com>. Accessed December 13, 2011.
3. Google. Google Scholar beta Web site. 2011; Available at <http://scholar.google.com>. Accessed December 13, 2011.
4. Jacques TS, Sebire NJ. The impact of article titles on citation hits: an analysis of general and specialist medical journals. *JRSM Short Rep.* 2010;1(1):2, <http://dx.doi.org/10.1258/shorts.2009.100020>.
5. Habibzadeh F, Yadollahie M. Are shorter article titles more attractive for citations? Cross-sectional study of 22 scientific journals. *Croat Med J.* 2010;51(2):165-70, <http://dx.doi.org/10.3325/cmj.2010.51.165>.
6. Neill US. How to write a scientific masterpiece. *J Clin Invest.* 2007;117(12):3599-602, <http://dx.doi.org/10.1172/JCI34288>.
7. Kliewer MA. Writing it up: a step-by-step guide to publication for beginning investigators. *AJR Am J Roentgenol.* 2005;185(3):591-6.
8. Vintzileos AM, Ananth CV. How to write and publish an original research article. *Am J Obstet Gynecol.* 2010;202(4):344.e1-6, <http://dx.doi.org/10.1016/j.ajog.2009.06.038>.
9. CLINICS. Instruction to Authors. 2012; Available at <http://www.clinics.org.br/inst.php>. Accessed February 24, 2012.
10. Archives of Internal Medicine. Manuscript criteria and information. 2011; Available at <http://archinte.ama-assn.org/misc/ifora.dtl#StructureofManuscript>. Accessed December 13, 2011.
11. British Medical Journal (BMJ). Title page. 2012; Available at <http://www.bmj.com/about-bmj/resources-authors/forms-policies-and-checklists/title-page>. Accessed January 18, 2012.
12. Siegel PZ, Thacker SB, Goodman RA, Gillespie C. Titles of Articles in Peer-Reviewed Journals Lack Information on Study Design: A Structured Review of Contributions to Four Leading Medical Journals, 1995 and 2001. *Sci Ed.* 2006;29(6):183-5.